



NC Public Schools – READY for Success

Educator Effectiveness and New Accountability Model Update

Spring 2012



Agenda

1. Why are you here?
2. Communication Plan
3. Educator Effectiveness Context and Policies
4. Case Study
5. New Accountability Model - Indicators and Reporting
6. Lunch
7. Measures of Student Learning
8. New Accountability Model - ESEA Waivers
9. How will you share with colleagues?



WHY ARE YOU HERE?



WHY ARE YOU HERE?



WHY ARE YOU HERE?



Communication Plan

How will you share the information learned from today with your colleagues?

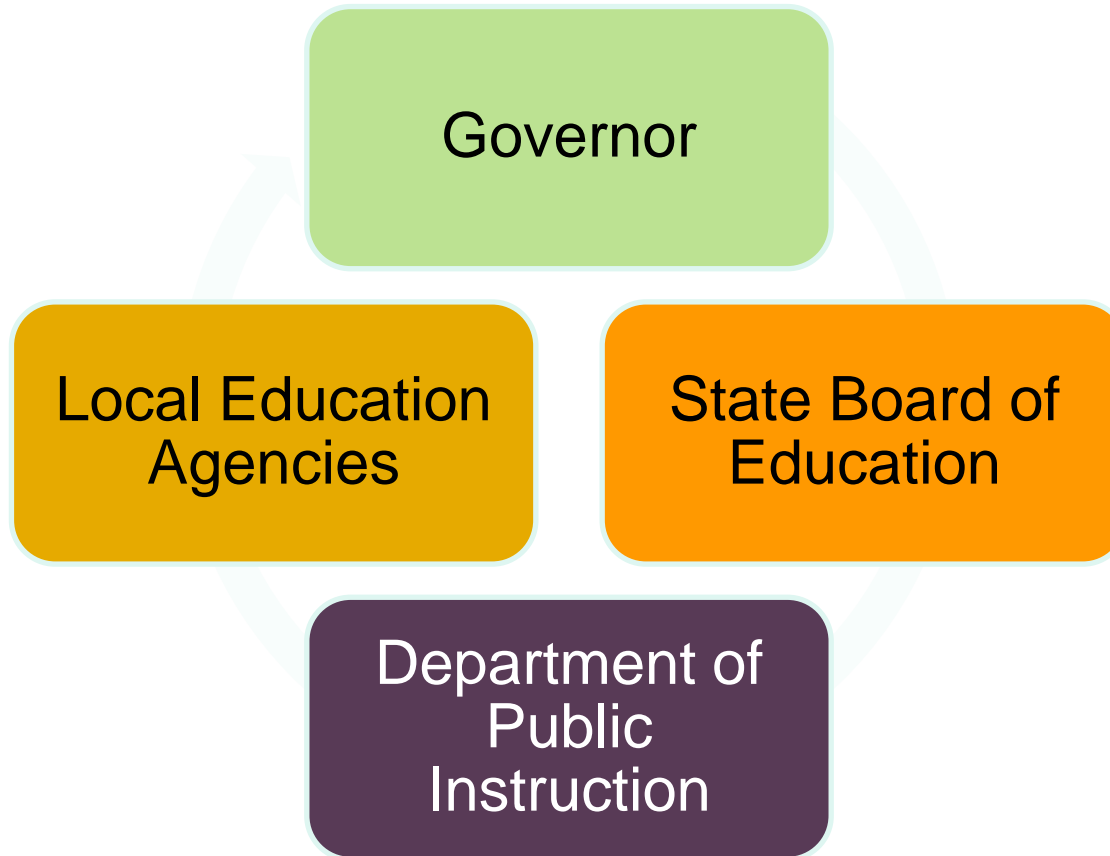
Discussion of communication plan handout



Context for Educator Effectiveness Work



Who are “We?”



What We Want:

Develop effective teachers and leaders in preparation programs



An effective teacher in every classroom and leader in every school



Achievement and growth for all students

Use professional development to increase effectiveness of teachers and leaders



What We Have

Statewide NC Educator Evaluation System

- All 115 school districts
- Many charter schools

Standard, Statewide Measures of Student Growth and Achievement

- English Language Arts
- Mathematics
- Some years of Science
- Career and Technical Education



What We Need

A required, standard, objective measurement of the effect each teacher has on each student's learning

Measures of student growth in all content areas

A common statewide process for identifying “effective” teachers and leaders



What We Need

A required, standard, objective measurement of the effect each teacher has on each student's learning

Measures of student growth in all content areas

A common statewide process for identifying “effective” teachers and leaders

Completed (Summer 2011)

Added Sixth and Eighth Standards

In Progress (2011-12)

Measures of Student Learning Design Process
(Phase One of Three Completed)

Targeted for Completion (February 2012)

Second Set of State Board Policies on Educator Effectiveness

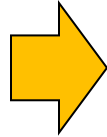


Today's Policies in Context of Overall Initiative



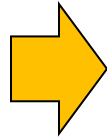
What We Need

Required, standard, objective measurement of the effect each teacher has on each student's learning



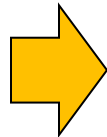
Completed (Summer 2011)
Added Sixth and Eighth Standards

Measures of student growth in all content areas



In Progress (2011-12)
Measures of Student Learning Design Process
(Phase One of Three Completed)

A common statewide process for identifying “effective” teachers and leaders



Targeted for Completion (February 2012)
Second Set of State Board Policies on Educator Effectiveness



What We Need

- Rating categories for the sixth and eighth standards
- Statewide growth model
- Statewide definitions of statuses of educators



Proposal for Revisions to TCP-C-006 Sixth and Eighth Standards



Key Elements

1. Yearly rating for teachers and administrators
2. Three rating categories
 - Does not meet expected growth
 - Meets expected growth
 - Exceeds expected growth

Student Growth Value

Administrators: average for all students

Teachers: three options for combining

- Students (of the teacher)
- School-wide (as above)



Three Options

The student growth value could be:

Option One

10% School-Wide

90% Individual

Option Two

20% School-Wide

80% Individual

Option
Three

30% School-Wide

70% Individual

Educator Effectiveness Work Group

The Work Group members were divided between Option One (90% individual; 10% school-wide) and Option Three (70% individual; 30% school-wide)



Option One

Pros:

- Bases an individual's evaluation heavily on an individual's student growth value
- Recognizes one teacher's ability to influence an entire faculty

Cons:

- Incentivizes competition between teachers
- Fails to recognize collaboration in PLCs



Option Three

Pros:

- **Incentivizes a school-wide sense of responsibility for the education of all students**
- **Recognizes the important role of PLCs, mentor teachers, and other support programs**

Cons:

- **Fails to recognize that one teacher may not be able to influence an entire faculty**
- **Makes an individual evaluation more dependent on a group outcome**



Proposal for Revisions to TCP-C-006 Statewide Growth Model



The Overarching Question

Which student academic growth model should NC use to determine sixth and eighth standard ratings for teachers and principals?



Sub-Questions

1. Which student academic growth models demonstrate technical adequacy?
2. Of those models that demonstrate technical adequacy, which meet other important criteria?



Value-Added Models

Use links between teachers and their students

Predict a student's score on an assessment

Attribute the difference between a predicted and actual score to above or below average instruction



Question One:

Which student academic growth models demonstrate technical adequacy?



How to Answer Question One

New, NC-specific study: Comparing Value-added Models For Estimating Teacher Effectiveness

Study completed by Drs. Rose, Henry, and Lauen of the University of North Carolina at Chapel Hill

Part of Race to the Top Evaluation



Answer to Question One

The **HLM3**, **URM**, and **SFE** are all technically adequate models to be used for determining sixth and eighth standard ratings for teachers and principals



Question Two:
**Of those models that
demonstrate technical
adequacy, which meet other
important criteria?**



How to Answer Question Two

Series of reports and recommendation by WestEd, a research and policy development non-profit agency selected by NCDPI to advise on selection of growth model

Dr. Carole Gallagher serves as lead consultant with support from technical and policy experts at WestEd



Research Process

Phase One: Conduct literature review

Phase Two: Identify evaluation criteria, including those other than technical adequacy

Phase Three: Review UNC-CH report

Phase Four: Make recommendation for SBE adoption of statewide model



Phase One

Phase One: Conduct literature review

Review of research literature that targeted theoretically and empirically based support for various statistical models of student growth

Identify the relative strengths and limitations of the models used most widely for the purpose of estimating teacher effectiveness.



Phase Two

Phase Two: Identify evaluation criteria, including those other than technical adequacy

Criteria include:

- Technical adequacy (validity, reliability, and fairness) of the model for the intended purpose
- Face validity with teachers and other stakeholders
- Theory- or research-based support
- Ease of use statewide for incorporating a measure of student growth in the NCEES
- Resource requirements
- Policy implications



Phase Three

Phase Three: Review UNC-CH report

WestEd technical experts agree with the methodology used in the study and praise the researchers for their extensive work

After reviewing the results of the various tests conducted, the experts agree with the three recommended models



Phase Four

Phase Four: Make recommendation for SBE adoption of statewide model

When evaluated with the six criteria, the **EVAAS URM** and **MRM** are the most well-suited for statewide use



Answer to Question Two

The **EVAAS URM** and **MRM** are technically strong, already known to educators within NC, supported by research, function within existing budget constraints, can be easily added to the NCEES, and are consistent with existing policy



The Overarching Question

Which student academic growth model should NC use to determine sixth and eighth standard ratings for teachers and principals?



Overall Recommendation

The SBE should adopt the EVAAS URM and MRM to determine sixth and eighth standard ratings for teachers and principals



Proposal for Revisions to TCP-C-006 Effectiveness Statuses



Three Statuses for Educators

1. In Need of Improvement

2. Effective

3. Highly Effective



Sixth/Eighth Standard

To determine effectiveness, a three-year rolling average of student growth values will be calculated:

Year One	Year Two	Year Three	Rolling Average
1.2	.8	1.0	1.0

And a rating will be assigned:

MEETS EXPECTED GROWTH



In Need of Improvement

Rating:

Standards 1-5

Any rating lower
than proficient

And/or

Standard 6

Does not meet
expected growth

Status:

**In Need of
Improvement**



In Need of Improvement

Rating:

Standards 1-7

Any rating lower
than proficient

And/or

Standard 8

“Does not meet
expected growth”

Status:

**In Need of
Improvement**



Effective

Rating of “Meets
expected growth”
on Standard Six

Proficient or
Higher on
Standards 1 - 5

Effective Teacher



Effective



Highly Effective

Rating of
“Exceeds
expected growth”
on Standard Six

Accomplished or
Higher on
Standards 1 - 5

Highly Effective
Teacher



Highly Effective



Sample Summary Sheet for Educator Evaluation



Display of All Information

- Ratings on Standards 1 – 5 of the Educator Evaluation System (as recorded in online tool).
- Yearly Standard 6 rating
- Three-year rolling average of student growth values and accompanying Standard 6 rating (for effectiveness determination)
- Overall teacher status



Summary Sheet

Draft January 2012: Assume SBE Adoption of Option One in TCP-C-006

North Carolina Educator Evaluation System Evaluation Summary Sheet

Name: Martha Washington **School:** Independence Elementary School
LEA: Freedom County Schools **Licensure:** Career-Status
Overall Status: Effective

Standard One: Teachers demonstrate leadership.														
Not Demonstrated			Developing			Proficient			Accomplished			Distinguished		
Standard Two: Teachers establish a respectful environment.														
Not Demonstrated			Developing			Proficient			Accomplished			Distinguished		
Standard Three: Teachers know the content they teach.														
Not Demonstrated			Developing			Proficient			Accomplished			Distinguished		
Standard Four: Teachers facilitate learning for their students.														
Not Demonstrated			Developing			Proficient			Accomplished			Distinguished		
Standard Five: Teachers reflect on their practice.														
Not Demonstrated			Developing			Proficient			Accomplished			Distinguished		
Standard Six: Teachers contribute to the academic success of students. * Only three-year rolling average is used to determine overall status *														
Year One (2009 – 2010)				Year Two (2010 – 2011)				Year Three (2011 – 2012)				Three-Year Rolling Average *		
Individual Student Growth: .8 School-wide Student Growth: .1 Year One Growth: .73				Individual Student Growth: 1.2 School-wide Student Growth: .5 Year Two Growth: 1.13				Individual Student Growth: .7 School-wide Student Growth: .5 Year Three Growth: .68				.85		
Does not meet expected growth	Meets expected growth	Exceeds expected growth		Does not meet expected growth	Meets expected growth	Exceeds expected growth		Does not meet expected growth	Meets expected growth	Exceeds expected growth		Does not meet expected growth	Meets expected growth	Exceeds expected growth
Overall Status:		Needs improvement				Effective				Highly Effective				



Case Study



New Accountability Model



Lunch



Measures of Student Learning (MSL) Quick Facts



Overview

1. **Measures of Student Learning (MSL) Quick Facts**
2. **MSL Design Process**
3. **Enthusiasm, Concerns, and Other Teacher Feedback**
4. **Challenges Ahead**
5. **Next Steps for 2011-12**



Guiding Principles

- North Carolina's experienced teachers know their students and their content
- They are the best-qualified to provide input on meaningful assessment of currently non-tested grades and subjects
- Valid measures of what students know and are able to do will likely exceed traditional multiple-choice assessment



What MSLs Are

- Measures of what students know and are able to do after completing a course or grade
- Tightly linked to the instruction that a teacher delivers
- **One** part of how North Carolina will evaluate the effectiveness of its teachers
- Similar to the common summative assessments that many districts already have in place



What MSLs Are Not

- Multiple-choice standardized exams for all areas of the Standard Course of Study
- Assessments that need to be delivered with the same level of security as EOCs and EOGs
- Designed without teacher input
- The only source of data used to make decisions about a teacher's effectiveness
- Part of the school accountability model



The Balancing Act

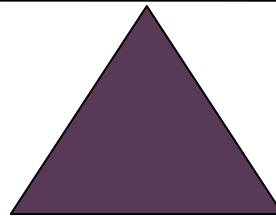
Given limited resources to dedicate to this work:

Freedom from Bias in
Results

High Levels of Reliability

Ability to Feed Results into
EVAAS or Another
Sophisticated Growth
Model

High Levels of Content
Validity, with Performance-
based Tasks



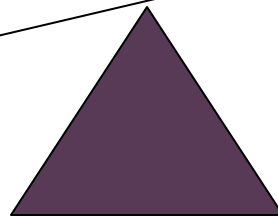
The Balancing Act

Freedom from Bias in Results

High Levels of Reliability

Ability to Feed Results into EVAAS or Another Sophisticated Growth Model

High Levels of Content Validity, with Performance-based Tasks



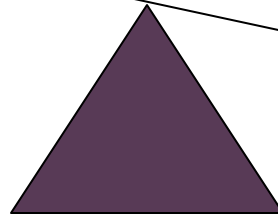
The Balancing Act

Freedom from Bias in
Results

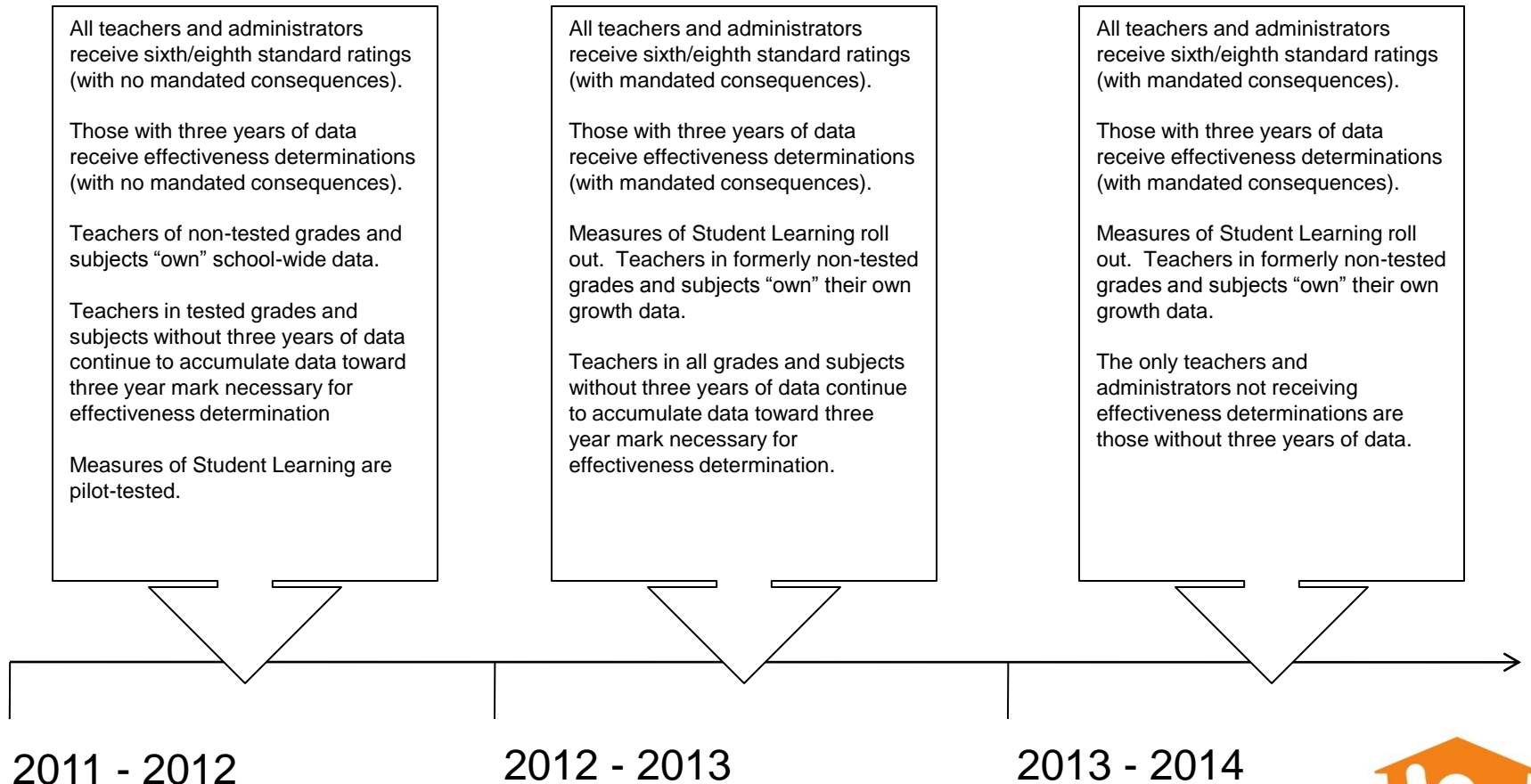
High Levels of Reliability

Ability to Feed Results into
EVAAS or Another
Sophisticated Growth
Model

High Levels of Content
Validity, with Performance-
based Tasks



Implementation Timeline



Four “Buckets” of Assessment

1

EOCs, EOGs
and VoCATS

2

Category One
of MSLs

With appropriate
resources and time,
these MSLs can be
validated
psychometrically

3

Category Two of
MSLs

With the heavy
emphasis on
performance, these
MSLs cannot be
validated
psychometrically

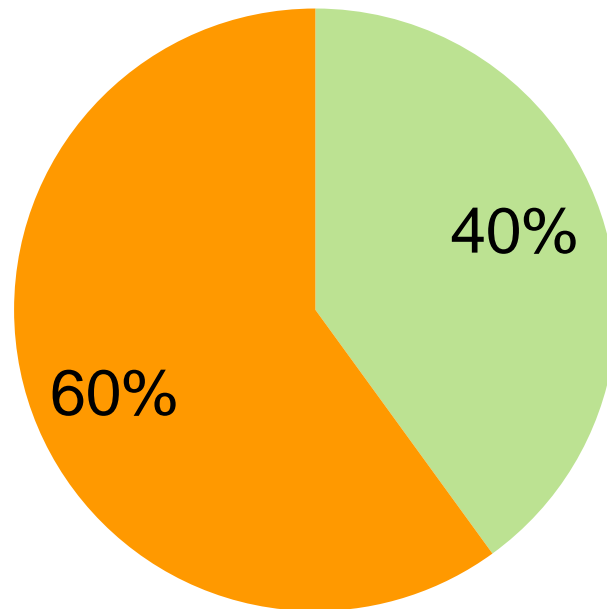
4

Locally
Developed
Courses



A Picture of Assessment

North Carolina Teacher Workforce



- EOCs, EOGs, or VoCATS
- Currently Non-Tested Grades and Subjects



Four “Buckets” of Assessment

1

EOCs, EOGs
and VoCATS

2

Category One
of MSLs

With appropriate
resources and time,
these MSLs can be
validated
psychometrically

3

Category Two of
MSLs

With their heavy
emphasis on
performance, these
MSLs cannot be
validated
psychometrically

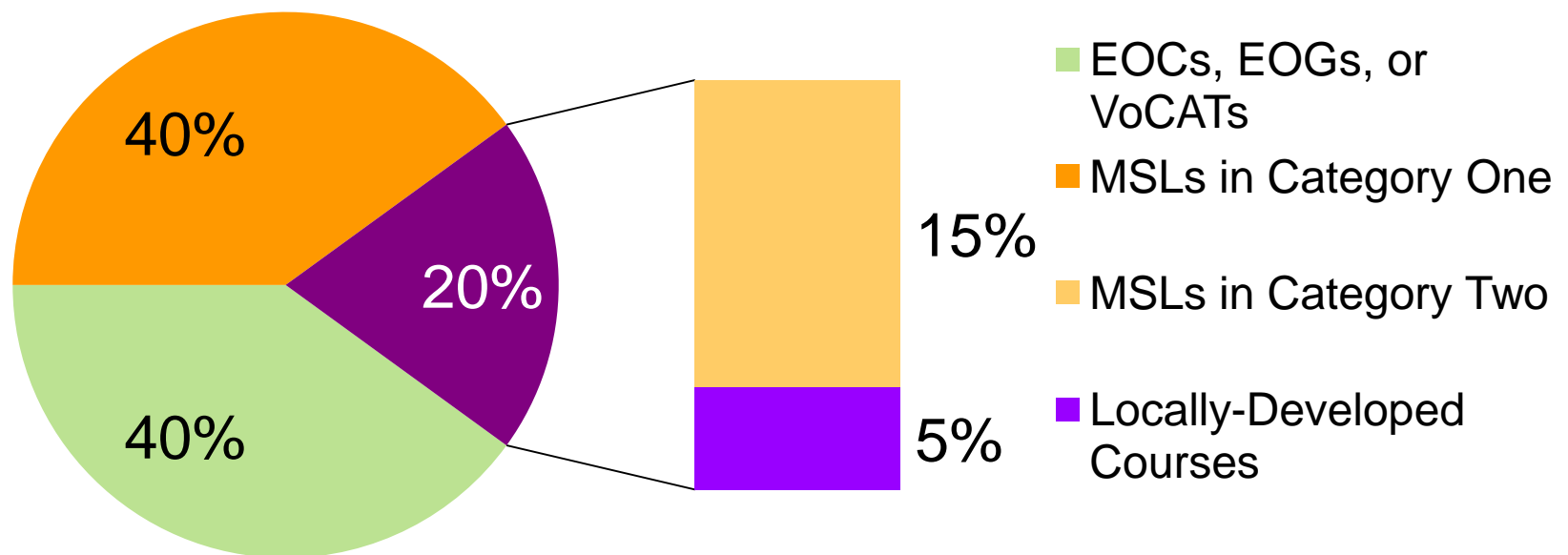
4

Locally
Developed
Courses



A Picture of Assessment

North Carolina Teacher Workforce



MSL Design Process



Three Phase Process

October 2011:

Teachers design item specifications for all currently non-tested grades and subjects

April – May 2012:

Teachers create rubrics and guidance for administering and scoring Measures of Student Learning

March 2012:

Teachers vet open-source items and items generated by external vendor(s)



Three Phase Process

Phase I

(October 2011)

Teachers design item specifications for all currently non-tested grades and subjects.

Phase III

(April – May 2012)

Teachers create rubrics and guidance for administering and scoring Measures of Student Learning.

Phase II

(March 2012)

Teachers evaluate open-source items and items generated by external vendor(s).



Phase I: Create Groups

**Developed list of all non-tested courses and grades in the
Common Core State Standards and NC Essential Standards**

Grouped courses and grades together into like-content groups

**Groups range from Extended Content Standards to Chemistry to
Elementary Theatre Arts to Social Studies Electives**



Phase I: Select Members

Designed an online application system for interested educators to apply to join one of the design groups

Advertised for the design groups

Selected educators from over 1,500 applications

Responded to LEA concerns and notified teachers of final selection decisions



Phase I: Ensure Representation

101 Local Education Agencies

10 Charter Schools

**Department of Juvenile Justice and Delinquency
Prevention**

University of North Carolina

North Carolina Virtual Public School



Phase I: Provide Training

1. **The Measures of Student Learning design process and how the Measures fit into the State's educator effectiveness work (DPI Leadership and RttT Project Management)**
2. **Assessment design, including potential item types, reliability, and validity (Assessment Design and Development)**
3. **Overview of the Common Core and NC Essential Standards (Curriculum & Instruction)**



Phase I: Gather Feedback

Through three feedback protocols, teachers provided answers to the following critical question:

**What does meaningful
assessment in your
content area look like?**



Preliminary Feedback:

To date, staff have consolidated input from K-2 English Language Arts and the Arts:

Subject Area	Teacher Input
K-2 ELA	<ul style="list-style-type: none">•Student growth needs to be measured throughout the year, as opposed to a “one-time” opportunity•Student growth is measured through student writing, speaking, listening, and reading•Kindergarten ELA will need to have some type of pre-assessment•In grades 1 – 2, the previous year’s score could count as the starting point for measuring growth•Reading 3-D is too time-intensive and used for formative purposes



Preliminary Feedback:

Subject Area	Teacher Input
The Arts	<ul style="list-style-type: none">•Growth in the arts can be measured through performance-based tasks and selected-response items•Performance-based tasks can assess the “creation” elements of the Essential Standards•Selected-response items can assess the “culture and history” elements of the Essential Standards•Performance-based tasks need to be broad enough to cover curricular options. For example, a scale can be sung, played on a string instrument, or played on a wind instrument



Teacher Thoughts and Concerns:

Appreciation, pride, and even joy that their content areas are now being valued and that the State is recognizing that they impact the learning of their students

Worry about finding a valid way to measure student learning in an art or PE class that meets once a month and one that meets every day

Value the input of teachers into the process

Doubt over the sustainability of these Measures of Student Learning after Race to the Top ends

Worry about a “test-heavy” environment for students, especially young children



Challenges Moving Forward



Resources:

Originally, North Carolina budgeted \$120,000 for this effort

With a Race to the Top amendment, North Carolina has \$1.6 million to pay for a vendor to generate items and \$720,000 to involve more teachers more frequently

\$1.6 million may not be enough funding to generate the robust item set described by the teacher design groups



Bias and Reliability in Grading:

Given the variety of items desired by the teacher design groups, educators will need to play a role in assessing student performance on Measures of Student Learning according to standardized rubrics

Their involvement introduces the risk for bias, even if teachers are not grading their own students' work

Teachers are concerned about when they will have time to grade the performance tasks in the Measures of Student Learning



Equality Among Content Areas:

Balance of instructional time between content areas

Ratings generated by a mathematical model and those selected by principals based on data



Next Steps for 2011 - 2012



Next Steps: 2011-12

- **Late December 2011/Early January 2012: Engage vendor to generate items**
- **March 2012: Design groups begin to return to vet items created by vendor**



Next Steps: 2011-12

- **April/May 2012:** Design groups return to design rubrics, administration instructions, and guidance on grading the Measures of Student Learning
- **Late Spring/Early Fall:** Pilot-test Measures of Student Learning



New Accountability Model

- *New Accountability Model(1 Hour, Thirty Minutes)*
- Display of the new accountability model



DPI Contacts

Educator Effectiveness

Dr. Rebecca Garland, Chief Academic Officer
Rebecca.Garland@dpi.nc.gov

Dr. Lynne Johnson, Director of Educator Recruitment and Development
Lynne.Johnson@dpi.nc.gov

Jennifer Preston, Race to the Top Project Coordinator
Jennifer.Preston@dpi.nc.gov

Dr. Eliz Colbert, Lead Professional Development Consultant
Eliz.Colbert@dpi.nc.gov

Robert Sox, Professional Development Consultant
Robert.Sox@dpi.nc.gov

DPI Contacts

New Accountability Model

Angela Quick, Deputy Chief Academic Officer
Angela.Quick@dpi.nc.gov

Dr. Tammy Howard, Director of Accountability Services
Tammy.Howard@dpi.nc.gov

Mike Martin, Policy Analyst
Mike.Martin@dpi.nc.gov



Design Your Message for Your Colleagues

ELIZ AND ROBERT ADD TEXT HERE

